



1309.43598X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: H. SUZUKI, et al

Serial No.: 10/790,140

Filing Date: March 2, 2004

For: DISK ARRAY DEVICE AND METHOD OF CHANGING THE
CONFIGURATION OF THE DISK ARRAY DEVICE

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

May 27, 2004

Sir:


Under the provisions of 35 USC 119 and 37 CFR 1.55, applicants hereby claim
the right of priority based on:

**Japanese Application No. 2004-000135
Filed: January 5, 2004**

A Certified copy of said application document is attached hereto.

Acknowledgement thereof is respectfully requested.

Respectfully submitted,



Carl I. Brundidge
Registration No. 29,621
ANTONELLI, TERRY, STOUT & KRAUS, LLP

CIB/jdc
Enclosures
703/312-6600

日本国特許庁
JAPAN PATENT OFFICE

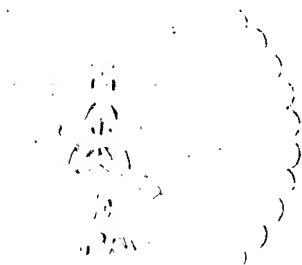
別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2004年 1月 5日
Date of Application:

出願番号 特願2004-000135
Application Number:
[ST. 10/C]: [JP 2004-000135]

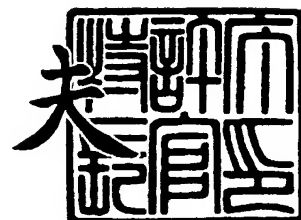
出願人 株式会社日立製作所
Applicant(s):



2004年 3月10日

特許庁長官
Commissioner,
Japan Patent Office

今井 康夫



出証番号 出証特2004-3018557

【書類名】 特許願
【整理番号】 340301729
【あて先】 特許庁長官殿
【国際特許分類】 G06F 03/00
G06F 01/18

【発明者】
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
D システム事業部内
【氏名】 鈴木 弘志

【発明者】
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
D システム事業部内
【氏名】 松重 博実

【発明者】
【住所又は居所】 神奈川県足柄上郡中井町境 7 8 1 番地 日立コンピューター機器
株式会社内
【氏名】 小川 正人

【発明者】
【住所又は居所】 神奈川県横浜市戸塚区吉田町 2 9 2 番地 株式会社日立アドバン
ストデジタル内
【氏名】 横山 智一

【特許出願人】
【識別番号】 000005108
【氏名又は名称】 株式会社日立製作所

【代理人】
【識別番号】 100095371
【弁理士】
【氏名又は名称】 上村 輝之

【選任した代理人】
【識別番号】 100089277
【弁理士】
【氏名又は名称】 宮川 長夫

【選任した代理人】
【識別番号】 100104891
【弁理士】
【氏名又は名称】 中村 猛

【手数料の表示】
【予納台帳番号】 043557
【納付金額】 21,000円

【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【包括委任状番号】 0110323

【書類名】 特許請求の範囲

【請求項1】

上位装置とのデータ授受を制御するチャネルアダプタと、
データを記憶する記憶デバイスと、
前記記憶デバイスが接続される記憶デバイス制御基板と、
前記記憶デバイス制御基板を介して前記記憶デバイスに接続され、前記記憶デバイスとのデータ授受を制御するディスクアダプタと、
前記ディスクアダプタ及び前記チャネルアダプタにそれぞれ接続された管理部と、を備え、

前記記憶デバイス制御基板は、前記記憶デバイスに接続される接続回路と、この接続回路の入力側及び出力側にそれぞれ設けられ、隣接する他の記憶デバイス制御基板と接続される連結モードと前記隣接する他の記憶デバイス制御基板と切り離される独立モードとを切替可能な切替回路とを含んで構成されており、

前記管理部からの出力信号によって前記切替回路の前記連結モードと前記独立モードとを切替可能に構成したディスクアレイ装置。

【請求項2】

前記記憶デバイス制御基板と前記他の記憶デバイス制御基板とは、それぞれ同一の取付用基板に実装されている請求項1に記載のディスクアレイ装置。

【請求項3】

前記切替回路が前記連結モードである場合、前記記憶デバイス制御基板と前記他の記憶デバイス制御基板とは、それぞれ同一の前記ディスクアダプタに接続されるようになっており、

前記切替回路が前記独立モードである場合、前記記憶デバイス制御基板と前記他の記憶デバイス制御基板とは、それぞれ別々の前記ディスクアダプタに接続されるようになっている請求項1に記載のディスクアレイ装置。

【請求項4】

前記記憶デバイスは第1ポート及び第2ポートを有し、前記第1ポート及び前記第2ポートは、それぞれ別々の前記記憶デバイス制御基板に接続され、かつ、これら各記憶デバイス制御基板は、それぞれ別々の前記ディスクアダプタに接続される請求項1に記載のディスクアレイ装置。

【請求項5】

前記接続回路は、ポートバイパスサーキットまたはファイバチャネルスイッチのいずれかにより構成される請求項1に記載のディスクアレイ装置。

【請求項6】

前記ディスクアダプタ及び前記記憶デバイス制御基板がそれぞれ備えるコネクタのうち入力側コネクタと出力側コネクタとにそれぞれ別々の色彩を関連付け、

前記各コネクタ間を接続する信号線のうち前記第1ポートに関連する信号線と前記第2ポートに関連する信号線とにそれぞれ別々の色彩を関連付けた請求項4に記載のディスクアレイ装置。

【請求項7】

上位装置とのデータ授受を制御するチャネルアダプタと、
データをそれぞれ記憶する複数の記憶デバイスと、
前記記憶デバイスとのデータ授受を制御するディスクアダプタと、
前記ディスクアダプタ及び前記チャネルアダプタにそれぞれ接続された管理部と、を備えたディスクアレイ装置の構成変更方法であって、

前記管理部から連結モードの指示が出された場合は、前記各記憶デバイスを連結させて同一の前記ディスクアダプタと接続させ、

前記管理部から独立モードの指示が出された場合は、前記各記憶デバイスを複数の記憶デバイス群に分割し、これら各記憶デバイス群をそれぞれ別々の前記ディスクアダプタと接続させる、ディスクアレイ装置の構成変更方法。

【請求項8】

取付用基板と、

前記取付用基板にそれぞれ設けられ、記憶デバイスに接続される複数の記憶デバイス制御基板とを備え、

前記記憶デバイス制御基板は、前記記憶デバイスに接続される接続回路と、前記接続回路の入力側に設けられた第1の切替回路と、前記接続回路の出力側に設けられた第2の切替回路とを含んで構成され、

外部から連結モード信号が入力された場合は、前記各切替回路を介して前記各記憶デバイス制御基板を接続して使用可能とし、外部から独立モード信号が入力された場合は、前記各記憶デバイス制御基板をそれぞれ分離して使用可能とした、記憶装置。

【書類名】明細書**【発明の名称】** ディスクアレイ装置及びディスクアレイ装置の構成変更方法**【技術分野】****【0001】**

本発明は、ディスクアレイ装置及びディスクアレイ装置の構成変更方法に関する。

【背景技術】**【0002】**

ディスクアレイ装置は、例えば、多数のディスクドライブをアレイ状に配設し、RAID (Redundant Array of Independent Inexpensive Disks) に基づいて構築されている。各ディスクドライブが有する物理的な記憶領域上には、論理的な記憶領域である論理ボリュームが形成されている。この論理ボリュームにはLUN (Logical Unit Number) が予め対応付けられている。ホストコンピュータは、LUN等を特定することにより、ディスクアレイ装置に対して所定形式の書込みコマンド又は読出しコマンドを発行する。これにより、ホストコンピュータは、ディスクアレイ装置に対して所望のデータの読み書きを行うことができる。

【0003】

図14に示すように、従来技術(特許文献1)では、複数のディスクアダプタ500のそれぞれに複数のディスクドライブ510を接続する。また、各ディスクドライブ510は、それぞれ複数のポートを有し、これら複数のポートは、それぞれ別々の経路を介してディスクアダプタ500に接続されている。各ディスクドライブ510のAポートは、接続基板520Aを介してDKA500に接続され、各ディスクドライブ510のBポートは、接続基板520Bを介してDKA500に接続される。従って、AポートまたはBポートのいずれか一方の経路に障害が発生した場合でも、交代パスとなっている他方の経路を介してディスクドライブ510にアクセスすることができる。また、図14に示す例では、それぞれ異なるDKA500に接続された複数のディスクドライブ510により、例えば、RAID5に従うRAIDグループが構成されている。従って、同一のRAIDグループに属するいずれか1つのディスクドライブ510に障害が発生した場合でも、同一のRAIDグループに属する他のディスクドライブ510に記憶されたデータに基づいて、データを復旧することができる。

【0004】

別の従来技術としては、図15に示すものが考えられる。この従来技術では、2つの接続基板611、612を介して、各ディスク制御部601、602と各ディスクドライブ610とをそれぞれ接続している。各ディスクドライブ610は、それぞれAポートとBポートとを備えている。図中左側に示すA側ディスク制御部601は、A側接続基板611を介して、複数のディスクドライブ610のAポートにそれぞれ接続されている。図中右側に示すB側ディスク制御部602は、B側接続基板612を介して、複数のディスクドライブ610のBポートにそれぞれ接続されている。

【特許文献1】 特開平7-20994号公報

【発明の開示】**【発明が解決しようとする課題】****【0005】**

近年では、ディスクアレイ装置のさらなる大容量化高性能化が求められているが、図14に示す従来技術のように、DKA500に接続されるディスクドライブ510の数を増加させるほど、インターフェース部のプロトコル変換等に要する処理時間が増大し、データ転送速度も低下する。従って、ディスクドライブ510の接続数を単純に増加させると、ディスクアレイ装置を利用するホストコンピュータから見た場合の書込み速度及び読出し速度が低下してしまう。

【0006】

また、図14中に示すように、もしも部位F1及びF2の2カ所でそれぞれ障害が発生した場合を考える。障害復旧のために、接続基板520Aまたは接続基板520Bのいず

れかを交換すると、障害部位F1、F2につながるディスクドライブ510の交代パスが失われることになる。

【0007】

即ち、例えば、接続基板520Aを先に交換する場合、障害部位F2につながるディスクドライブ510は、Aポート及びBポートのいずれからもアクセスすることができなくなる。Aポート側の経路は、接続基板520Aを取り外した時点で失われ、Bポート側の経路は障害部位F2により使用不能だからである。接続基板520Bを先に交換する場合、障害部位F1につながるディスクドライブ510の交代パスが失われる。即ち、このディスクドライブ510のAポート側経路は障害部位F1により使用不能であり、Bポート側経路は接続基板520Bを取り外した時点で失われる。

【0008】

障害部位F1、F2と無関係のディスクドライブ510にはアクセス可能である。従って、障害部位F1、F2の両方で障害が発生した場合は、接続基板520A、520Bをそれぞれ正常品に交換した後で、アクセス不能となったディスクドライブ510に書き込むべきデータを、同一RAIDグループ内の他のディスクドライブ510の記憶内容に基づいて復元する。このデータ復元処理（データ回復処理）は、データ回復に関するディスクドライブ510に新たな障害が発生する前に完了しなければならない。データ回復処理の完了前に、さらに新たな障害が発生した場合は、データ回復を行うことができなくなるためである。RAID5では、同一RAIDグループ内のいずれか1つのディスクドライブにアクセス不能であっても、残りのディスクドライブの記憶内容に基づいてデータを回復可能である。しかし、RAID5では、同一RAIDグループ内で複数のディスクドライブにアクセス不能となった場合、データを回復することはできない。

【0009】

このように、複数の障害部位F1、F2でそれぞれ障害が発生した場合は、新たな障害が発生する前に、データ回復処理を終える必要がある。しかし、DKA500に接続されるディスクドライブ510の数が増大傾向にあるため、データ回復に要する期間も長期化する傾向がある。また、データ回復処理が完了するまで他のディスクドライブ510に新たな障害が発生するのを極力防止する必要がある。しかし、近年は、各種基板の高密度実装化、データ転送速度やドライブアクセス速度の高速化等が要求されるため、接続基板520A、520Bやディスクドライブ510の障害発生率を現在よりも大幅に低下させるのは簡単なことではない。

【0010】

一方、図15に示す従来技術では、A側接続基板611が複数のディスクドライブ群のAポートへの接続を担当し、B側接続基板612が複数のディスクドライブ群のBポートへの接続を担当する。従って、いずれか一方の接続基板に障害が発生して交換する場合、障害が発生した接続基板を取り外すと、複数のディスクドライブ群に影響を与えることになる。従って、図15に示す従来技術では、図14と共に述べた問題が発生する上に、影響範囲も増加する可能性がある。

【0011】

そこで、本発明の1つの目的は、障害発生に対する耐性を向上できるようにしたディスクアレイ装置及びディスクアレイ装置の構成変更方法を提供することにある。本発明の1つの目的は、共通化された構造によって複数の使用目的に対応できるようにしたディスクアレイ装置及びディスクアレイ装置の構成変更方法を提供することにある。本発明の1つの目的は、メンテナンス性及び信頼性を向上できるようにしたディスクアレイ装置及びディスクアレイ装置の構成変更方法を提供することにある。本発明の他の目的は、後述する実施の形態の記載から明らかになるであろう。

【課題を解決するための手段】

【0012】

上記課題を解決すべく、本発明に従うディスクアレイ装置は、上位装置とのデータ授受を制御するチャネルアダプタと、データを記憶する記憶デバイスと、記憶デバイスが接続

される記憶デバイス制御基板と、記憶デバイス制御基板を介して記憶デバイスに接続され、記憶デバイスとのデータ授受を制御するディスクアダプタと、ディスクアダプタ及びチャネルアダプタにそれぞれ接続された管理部と、を備えている。そして、記憶デバイス制御基板は、記憶デバイスに接続される接続回路と、この接続回路の入力側及び出力側にそれぞれ設けられ、隣接する他の記憶デバイス制御基板と接続される連結モードと、隣接する他の記憶デバイス制御基板と切り離される独立モードとを切替可能な切替回路とを含んで構成されている。さらに、管理部からの出力信号によって、切替回路の連結モードと独立モードとを切替可能に構成している。

【0013】

管理部から連結モードが指示されると、切替回路によって一方の記憶デバイス制御基板と他方の記憶デバイス制御基板とが連結される。また、管理部から独立モードが指示されると、切替回路によって一方の記憶デバイス制御基板と他方の記憶デバイス制御基板とがそれぞれ分離する。従って、同一の基本構造でありながら、連結モードではより多くの記憶デバイスをディスクアダプタに接続することができ、独立モードでは、より多くのディスクアダプタにより記憶デバイスを制御することができる。これにより、ユーザの使用目的に応じた構成を比較的容易に実現することができる。

【0014】

本発明の一態様では、記憶デバイス制御基板と他の記憶デバイス制御基板とは、それぞれ同一の取付用基板に実装されている。

【0015】

本発明の一態様では、切替回路が連結モードである場合、記憶デバイス制御基板と他の記憶デバイス制御基板とは、それぞれ同一のディスクアダプタに接続されるようになっており、切替回路が独立モードである場合、記憶デバイス制御基板と他の記憶デバイス制御基板とは、それぞれ別々のディスクアダプタに接続されるようになっている。

【0016】

本発明の一態様では、記憶デバイスは第1ポート及び第2ポートを有し、第1ポート及び第2ポートは、それぞれ別々の記憶デバイス制御基板に接続され、かつ、これら各記憶デバイス制御基板は、それぞれ別々のディスクアダプタに接続されている。

【0017】

本発明の一態様では、接続回路は、ポートバイパスサーキットまたはファイバチャネルスイッチのいずれかにより構成されている。

【0018】

本発明の一態様では、ディスクアダプタ及び記憶デバイス制御基板がそれぞれ備えるコネクタのうち入力側コネクタと出力側コネクタとにそれぞれ別々の色彩を関連付け、各コネクタ間を接続する信号線のうち第1ポートに関連する信号線と第2ポートに関連する信号線とにそれぞれ別々の色彩を関連付けている。

【発明を実施するための最良の形態】

【0019】

以下、図1～図13に基づき、本発明の実施の形態を説明する。本実施形態では、記憶デバイスに接続される接続回路と、この接続回路の入力側及び出力側にそれぞれ設けられ、隣接する他の記憶デバイス制御基板と接続される連結モードと隣接する他の記憶デバイス制御基板と切り離される独立モードとを切替可能な切替回路とを含んで構成される記憶デバイス制御基板が開示されている。そして、管理部からの出力信号によって切替回路の連結モードと独立モードとが切替可能となっている。

【0020】

また、本実施形態では、上位装置とのデータ授受を制御するチャネルアダプタと、データをそれぞれ記憶する複数の記憶デバイスと、記憶デバイスとのデータ授受を制御するディスクアダプタと、ディスクアダプタ及びチャネルアダプタにそれぞれ接続された管理部と、を備えたディスクアレイ装置の構成変更方法が開示されている。そして、この構成変更方法では、管理部から連結モードの指示が出された場合は、各記憶デバイスを連結させ

て同一の前記ディスクアダプタと接続させ、管理部から独立モードの指示が出された場合は、各記憶デバイスを複数の記憶デバイス群に分割し、これら各記憶デバイス群をそれぞれ別々のディスクアダプタと接続させる。

【実施例1】

【0021】

図1は、ディスクアレイ装置10の全体概要を示すブロック図である。ディスクアレイ装置10は、通信ネットワークCN1を介して、複数のホストコンピュータ1と双方向通信可能にそれぞれ接続されている。ここで、通信ネットワークCN1は、例えば、LAN (Local Area Network)、SAN (Storage Area Network)、インターネットあるいは専用回線等である。LANを用いる場合、ホストコンピュータ1とディスクアレイ装置10との間のデータ転送は、TCP/IP (Transmission Control Protocol/Internet Protocol) プロトコルに従って行われる。SANを用いる場合、ホストコンピュータ1とディスクアレイ装置10とは、ファイバチャネルプロトコルに従ってデータ転送を行う。また、ホストコンピュータ1がメインフレームの場合は、例えば、FICON (Fibre Connection: 登録商標)、ESCON (Enterprise System Connection: 登録商標)、ACONARC (Advanced Connection Architecture: 登録商標)、FIBARC (Fibre Connection Architecture: 登録商標) 等の通信プロトコルに従ってデータ転送が行われる。

【0022】

各ホストコンピュータ1は、例えば、サーバ、パーソナルコンピュータ、ワークステーション、メインフレーム等として実現されるものである。例えば、各ホストコンピュータ1は、図外に位置する複数のクライアント端末と別の通信ネットワークを介して接続されている。各ホストコンピュータ1は、例えば、各クライアント端末からの要求に応じて、ディスクアレイ装置10にデータの読み書きを行うことにより、各クライアント端末へのサービスを提供する。

【0023】

ディスクアレイ装置10は、それぞれ後述するように、複数のチャネルアダプタ (以下、CHAと略記) 20と、複数のディスクアダプタ (以下、DKAと略記) 30と、キャッシュメモリ40と、共有メモリ50と、スイッチ部60と、SVP70と、ディスク駆動部80とを備えている。また、ディスクアレイ装置10には、例えばLAN等の通信ネットワークCN2を介して、管理端末2が接続されている。

【0024】

ディスクアレイ装置10には、例えば4個、8個等のように複数のCHA20を設けることができる。各CHA20は、それぞれに接続されたホストコンピュータ1から、データの読み書きを要求するコマンド及びデータを受信し、ホストコンピュータ1から受信したコマンドに従って動作する。DKA30の動作も含めて先に説明すると、例えば、CHA20は、ホストコンピュータ1からデータの読出し要求を受信すると、読出しコマンドを共有メモリ50に記憶させる。DKA30は、共有メモリ50を随時参照しており、未処理の読出しコマンドを発見すると、ディスクドライブ81からデータを読み出して、キャッシュメモリ40に記憶させる。CHA20は、キャッシュメモリ40に移されたデータを読み出し、コマンド発行元のホストコンピュータ1に送信する。

【0025】

また例えば、CHA20は、ホストコンピュータ1からデータの書き込み要求を受信すると、書き込みコマンドを共有メモリ50に記憶させると共に、受信したデータ (ユーザデータ) をキャッシュメモリ40に記憶させる。CHA20は、キャッシュメモリ40にデータを記憶した後、ホストコンピュータ1に対して書き込み完了を報告する。そして、DKA30は、共有メモリ50に記憶された書き込みコマンドに従って、キャッシュメモリ40に記憶されたデータを読出し、所定のディスクドライブ81に記憶させる。

【0026】

ディスクアレイ装置10には、例えば4個や8個等のように複数のDKA30を設けることができる。各DKA30は、各ディスクドライブ81との間のデータ通信を制御する

もので、それぞれプロセッサ部と、データ通信部と、ローカルメモリ（いずれも不図示）と、FC制御部31（図3参照）等を備えている。各DKA30と各ディスクドライブ81とは、例えば、SAN等の通信ネットワークを介して接続されており、ファイバチャネルプロトコルに従ってブロック単位の水タ転送を行う。

【0027】

各DKA30は、ディスクドライブ81の状態を随時監視しており、この監視結果は内部の通信ネットワークCN3を介してSVP70に送信される。なお、各CHA20及び各DKA30は、例えば、プロセッサやメモリ等が実装されたプリント基板と、メモリに格納された制御プログラムとをそれぞれ備えており、これらのハードウェアとソフトウェアとの協働作業によって、所定の機能を実現する。

【0028】

キャッシュメモリ40は、例えば、ユーザデータ等を記憶するものである。キャッシュメモリ40は、例えば不揮発メモリから構成される。キャッシュメモリ40は、複数のメモリから構成することができ、ユーザデータを多重管理することができる。

【0029】

共有メモリ（あるいは制御メモリ）50は、例えば不揮発メモリから構成される。共有メモリ50には、例えば、制御情報等が記憶される。なお、制御情報等の情報は、複数の共有メモリ50により多重管理することができる。共有メモリ50及びキャッシュメモリ40は、それぞれ複数個設けることができる。

【0030】

スイッチ部60は、各CHA20と、各DKA30と、キャッシュメモリ40と、共有メモリ50とをそれぞれ接続するものである。これにより、全てのCHA20、DKA30は、キャッシュメモリ40及び共有メモリ50にそれぞれアクセス可能である。

【0031】

SVP（Service Processor）70は、内部LAN等の通信ネットワークCN3を介して、各CHA20及び各DKA30から情報を収集するものである。SVP70が収集する情報としては、例えば、装置構成、電源アラーム、温度アラーム、入出力速度（IOPS）等が挙げられる。SVP70は、通信ネットワークCN2を介して管理端末2に接続されている。管理端末2は、SVP70により収集された各種情報を閲覧等できる。また、管理端末2は、SVP70を介して、例えば、RAID設定や閉塞処理、後述する構成変更等を指示することができる。

【0032】

ディスクアレイ装置10は、少なくとも1つ以上のディスク駆動部80を備える。図示の例では、4つのディスク駆動部80を示してある。各ディスク駆動部80のバックボードには、それぞれ複数のディスクドライブ81が実装されている。各ディスクドライブ81は、例えば、ハードディスク装置や半導体メモリ装置等して実現可能である。複数のディスクドライブ81によりRAIDグループを構成することができ、このRAIDグループが提供する物理的な記憶領域上に、論理的な記憶領域（論理ボリューム（Logical Unit）あるいは論理デバイス（LDEV））を設定することができる。また、各ディスク駆動部80のバックボードには、各ディスクドライブ81に接続するためのHDD制御基板82が実装されている。

【0033】

HDD制御基板82は、ディスクドライブ81の各ポート側にそれぞれ複数個ずつ設けられている。即ち、例えば、各ディスクドライブ81の一方のポート側には2個のHDD制御基板82が設けられ、各ディスクドライブ81の他方のポート側にも2個のHDD制御基板82が設けられている。このように、本実施例では、複数のHDD制御基板82によって、ディスクドライブ群の各ポートへの経路を形成しており、この経路の構成を変更可能としている。そして、各HDD制御基板82は、ケーブル90を介して所定のDKA30とそれぞれ接続されている。

【0034】

図 2 は、ディスクアレイ装置 1 0 を正面から模式的に示す説明図である。ディスクアレイ装置 1 0 は、例えば、基本部 1 0 1 と、増設部 1 0 2 とから構成できる。基本部 1 0 1 は、ディスク制御部 1 1 と、ディスク駆動部 8 0 とを備えている。ディスク制御部 1 1 は、ディスクアレイ装置 1 0 の全体的な制御を行うもので、各 C H A 2 0、各 D K A 3 0、キャッシュメモリ 4 0、共有メモリ 5 0、スイッチ部 6 0 及び S V P 7 0 等を含んで構成することができる。増設部 1 0 2 は、複数のディスク駆動部 8 0 から構成できる。増設部 1 0 2 の制御は、基本部 1 0 1 のディスク制御部 1 1 により行われる。従って、ディスクアレイ装置 1 0 の最小構成は、基本部 1 0 1 のみとなる。増設部 1 0 2 は、必要に応じて追加可能なオプションである。

【 0 0 3 5 】

図 2 に示す例では、複数のディスクドライブ群が連結モードで接続され、大容量の記憶領域を実現している。増設部 1 0 2 の図中上段に並んで位置する 2 個のディスクドライブ群は、互いに連結されており、基本部 1 0 1 の左側に示すディスクドライブ群とケーブル 9 0 を介して接続されている。同様に、増設部 1 0 2 の下段に並んで位置する 2 個のディスクドライブ群も互いに連結されており、基本部 1 0 1 の右側に示すディスクドライブ群と別のケーブル 9 0 を介して接続されている。従って、図示の例では、3 個のディスクドライブ群を連結したディスクドライブグループが合計 2 個示されているが、これは説明の便宜上の例であって、実際には、より多くのディスクドライブグループを構成することができる。それぞれ別々の D K A 3 0 により制御することができる。

【 0 0 3 6 】

図 3 は、D K A 3 0 及びディスク駆動部 8 0 を中心とした論理的な全体構成を示す説明図である。図 3 に示す例では、2 個の D K A 3 0 (# 0、# 1) と複数のディスク駆動部 8 0 とが示されている。

【 0 0 3 7 】

各 D K A 3 0 には、ディスク駆動部 8 0 の数に一致する数だけ F C 制御部 3 1 がそれぞれ設けられている。F C 制御部 3 1 は、例えば、ファイバチャネルプロトコルへの変換処理等を行うもので、ディスクドライブ 8 1 とのデータ入出力を実際に制御する制御論理回路である。各 F C 制御部 3 1 は、通信ネットワーク C N 3 から S V P 7 0 を介して、管理端末 2 に接続されている。また、各 F C 制御部 3 1 は、スイッチ部 6 0 を介して、C H A 2 0 やキャッシュメモリ 4 0、共有メモリ 5 0 に接続される。

【 0 0 3 8 】

さらに、各 F C 制御部 3 1 は、ケーブル 9 0 を介して所定のディスク駆動部 8 0 にそれぞれ接続されている。各 F C 制御部 3 1 は、ディスク駆動部 8 0 が有する 2 種類のポート群のうち所定のポート群にのみ接続される。従って、各ディスク駆動部 8 0 には、2 つの F C 制御部 3 1 が接続される。これら 2 つの F C 制御部 3 1 は、それぞれ別の D K A 3 0 に属する。従って、もしもいずれか一方の D K A 3 0 に障害が発生した場合でも、他方の D K A 3 0 から交代パスを介して、ディスク駆動部 8 0 のディスクドライブ群にアクセスすることができる。

【 0 0 3 9 】

ディスク駆動部 8 0 には、複数のディスクドライブ 8 1 がバックボード（図示せず）に着脱可能に取り付けられている。図示の例では、# 0 ~ # n までの n + 1 個のディスクドライブ 8 1 によって 1 つのディスクドライブグループが形成されており、1 つのディスク駆動部 8 0 に 2 つのディスクドライブグループが設けられている。各ディスクドライブグループには、各ポート側にそれぞれ H D D 制御基板 8 2 が設けられている。各ディスクドライブ 8 1 は、自己に接続された 2 つの H D D 制御基板 8 2 のうちいずれか一方または双方を介して、D K A 3 0 の F C 制御部 3 1 とデータ入出力を行うことができる。

【 0 0 4 0 】

各 H D D 制御基板 8 2 は、接続回路 2 0 0 と、接続回路 2 0 0 の入力側及び出力側にそれぞれ接続された切替回路 2 1 0 とを備えている。図 4 は、H D D 制御基板 8 2 のより詳細な構造を示すブロック図である。

【0041】

各切替回路210は、それぞれ2個のスイッチ211、212から構成される。各スイッチ211、212は、例えば、単極双投型（SPDT）のスイッチ回路として構成可能である。各スイッチ211、212の接点bは、外部接続用の接点（以下「外部接点b」）であり、各スイッチ211、212の接点aは、内部接続用の接点（以下「内部接点a」）である。接点cは、共通接点である。

【0042】

図4に示す例において、入力側（図中左側）の切替回路210に着目すると、各スイッチ211、212の外部接点bには、ケーブル90を介してFC制御部31がそれぞれ接続されている。そして、各スイッチ211、212の外部接点bと共通接点cとがそれぞれ接続されることにより、FC制御部31がHDD制御基板82に接続されている。出力側（図中右側）の切替回路210に着目すると、各スイッチ211、212の内部接点aは、ディスク駆動部80のバックボードに形成されたプリント配線にそれぞれ接続されており、また、各スイッチ211、212の共通接点cは、内部接点aとそれぞれ接続されている。これにより、隣接するHDD制御基板82同士は、バックボードのプリント配線及び切替回路210を介して連結される。なお、出力側の各スイッチ211、212の外部接点bはそれぞれ開放されているが、例えば、ジャンパー線等によって各外部接点b同士を接続してもよい。

【0043】

ここで、各切替回路210の各スイッチ211、212は、ディスク駆動部80に設けられたコネクタ83に接続されており、コネクタ83を介してFC制御部31や隣接する他のHDD制御基板82と接続される。

【0044】

接続回路200及びFC制御部31には、それぞれSERDES（Serializer and Deserializer）が設けられている。SERDESとは、シリアルデータをパラレルデータに、またパラレルデータをシリアルデータに、それぞれ変換する変換回路である。接続回路200内には、その入力側にSERDES201が設けられ、その出力側にSERDES202がそれぞれ設けられている。入力側SERDES201と出力側SERDES202との間は、内部バス204によってパラレル接続されている。また、内部バス204には、ディスク側SERDES203が複数接続されている。ディスク側SERDES203は、HDD制御基板82が管理する各ディスクドライブ81毎にそれぞれ1つずつ設けられる。

【0045】

入力側SERDES201は、HDD制御基板82の外部から入力されたシリアルデータをパラレルデータに変換し、内部バス204に送出する。出力側SERDES202は、内部バス204を介して受信したパラレルデータをシリアルデータに変換し、HDD制御基板82の外部に送信する。ディスク側SERDES203は、内部バス204を介して受信したパラレルデータをシリアルデータに変換し、ディスクドライブ81に書込む。あるいは、ディスク側SERDES203は、ディスクドライブ81から読み出したシリアルデータをパラレルデータに変換して内部バス204に送出する。各ディスク側SERDES203は、内部バスを介して受信したパラレルデータが自己宛（自己が担当するディスクドライブ81宛）のデータであるか否かを判定し、自己宛のデータである場合に作動して、ディスクドライブ81への入出力を行う。自己宛のデータであるか否かは、例えば、受信したデータ中に含まれるディスクドライブ番号等に基づいて判断できる。

【0046】

図5は、隣接するディスクドライブグループを連結した場合、即ち、隣接するHDD制御基板82（#0、#1）を接続する場合を示すブロック図である。図5中では、説明の便宜上、各HDD制御基板82についてそれぞれ1つだけディスクドライブ81を示しているが、実際には、複数のディスクドライブ81が接続される。

【0047】

各HDD制御基板82は、それぞれ同一の構成を有する。前段のHDD制御基板82（

#0)の出力側と後段のHDD制御基板82(#1)の入力側とは、ディスク駆動部80のバックボードに形成されたプリント配線を介して接続されている。前段のHDD制御基板82(#0)に着目すると、入力側のスイッチ211, 212は、FC制御部31のSERDES32にそれぞれ接続された外部接点bと共通接点cとが接続されており、出力側のスイッチ211, 212は、共通接点cと内部接点aとがそれぞれ接続されている。従って、前段のHDD制御基板82(#0)は、その入力側が外部のFC制御部31とケーブル90を介して接続されると共に、その出力側がバックボードのプリント配線を介して、隣接する後段のHDD制御基板82(#1)に接続されている。

【0048】

後段のHDD制御基板82(#1)に着目すると、入力側及び出力側のスイッチ211, 212は、それぞれ内部接点aと共通接点cとが接続されている。また、出力側のスイッチ211, 212は、内部接点a同士が例えばジャンパー線等の導線213を介して接続されている。従って、後段のHDD制御基板82(#1)は、プリント配線と内部接点a及び共通接点cとを介して、前段のHDD制御基板82(#0)に縦続接続される。

【0049】

信号の伝達経路について説明する。FC制御部31からケーブル90及びコネクタ83を介して前段のHDD制御基板82(#0)に入力されたシリアルデータは、入力側スイッチ211の外部接点bから共通接点cを介して入力側SERDES201に入力され、このSERDES201によりパラレルデータに変換される。このパラレルデータは、内部バス204を介して出力側SERDES202に入力され、シリアルデータに変換される。このシリアルデータは、出力側のスイッチ211の共通接点cから内部接点aを介して、ディスク駆動部80のバックボードに形成されたプリント配線に送出される。

【0050】

プリント配線に送出されたシリアルデータは、後段のHDD制御基板82(#1)のコネクタ83から入力側スイッチ211の内部接点a及び共通接点cを介して、入力側SERDES201に入力される。そして、SERDES201によりパラレルデータに変換されて内部バス204に送出され、出力側のSERDES202に到達する。出力側SERDES202によりシリアルデータに変換されたデータは、出力側のスイッチ211の共通接点c及び内部接点aから導線213を介して、他方の出力側スイッチ212に入力される。データは、上述の経路を逆に通って、前段のHDD制御基板82(#0)に戻り、HDD制御基板82(#0)からケーブル90を介して、FC制御部31に戻る。

【0051】

このように、図5に示す連結モードでは、ディスク駆動部80に実装された複数のディスクドライブグループが1つに連結されて大きな単一のグループを構成する。図6は、連結モードにおいて障害が発生した場合の様子を示す説明図である。例えば、上側に示すディスク駆動部80内で、同時に2カ所で障害が発生したとする(左側の#2のディスクドライブに接続される経路と、右側の#nのディスクドライブに接続される経路)。この場合、各ディスクドライブグループ毎にHDD制御基板82が設けられているため、換言すれば、一連のディスクドライブ群に接続するための回路が複数のHDD制御基板82として分割されているため、障害に関連するHDD制御基板82のみを交換可能である。従って、障害への耐性が向上し、信頼性が高まる。

【0052】

次に、図7～図11に基づいて、独立モードの構成を説明する。図1は、ディスクアレイ装置10の全体構成を示す説明図である。ディスクアレイ装置10を独立モードで運用する場合、システム管理者が管理端末2を介して、構成変更を指示する。この指示を受けて、切替回路210が切り替わり、連結モードから独立モードへ移行可能となる。ディスクアレイ装置10は、連結モードまたは独立モードのいずれかで運用することができ、いずれの形態を採用するかは、例えばシステム管理者により決定される。本実施形態では、切替回路210の切替動作等だけで独立モードと連結モードとを相互に切替可能であり、多くの構成が共通する。従って、既に述べた構成と重複する説明は省略し、独立モードに

特有の構成を中心に説明する。

【0053】

独立モードでは、同一のディスク駆動部80に実装された複数のディスクドライブ群（ディスクドライブグループ）が、それぞれ別々に使用される。図示の例では、各ディスク駆動部80に2つのディスクドライブ群を実装しているため、個別に運用されるディスクドライブ群の数は、連結モードの場合と比べて2倍となる。従って、独立モードの場合、ディスク制御部11にDKA30をさらに2個追加する。

【0054】

図8の正面概略図に示すように、増設部102の各ディスク駆動部80は、それぞれ2個ずつのディスクドライブ群を有する。これら各ディスクドライブ群への接続を担当するHDD制御基板82には、それぞれ個別にケーブル90が接続されており、同一のディスク駆動部80に実装された2つのディスクドライブ群はそれぞれ個別に運用されるようになっている。

【0055】

図9は、独立モード運用時におけるDKA30及びディスク駆動部80等の構成を示す説明図である。図3に示す連結モードの構成と比較すると、図9に示す独立モードの構成では、各ディスク駆動部80内において、各ディスクドライブ群はそれぞれ切り離されている。そして、これら各ディスクドライブ群を担当するHDD制御基板82は、それぞれケーブル90を介して、所定のDKA30と接続されている。

【0056】

図10は、同一のディスク駆動部80において、隣接するHDD制御基板82をそれぞれ別々のDKA30（FC制御部31）に接続する様子を示すブロック図である。独立モードの場合、HDD制御基板82（#0）とHDD制御基板82（#1）とは互いに接続されず、分離される。そして、一方のHDD制御基板82（#0）には、一方のDKA30のFC制御部31（#0）がケーブル90を介して接続される。また、他方のHDD制御基板82（#1）には、他方のDKA30のFC制御部31（#1）が別のケーブル90を介して接続される。従って、各HDD制御基板82（#0、#1）がそれぞれ担当するディスクドライブ群は、それぞれ別々のFC制御部31（#0、#1）によりデータの入出力が行われる。これにより、ディスクアレイ装置10の全体的な性能は、連結モードに比べて高くなる。従って、例えば、独立モードを高容量及び高性能モードと、連結モードを高容量及び低性能モードと表現可能である。

【0057】

図10に示すように、独立モードの場合は、HDD制御基板82（#0）の出力側スイッチ211、212の内部接点a同士が導線213で接続されており、また、HDD制御基板82（#1）の出力側スイッチ211、212の内部接点a同士も別の導線213で接続されている。

【0058】

従って、独立モードの場合の信号伝達経路を説明すると、FC制御部31からケーブル90を介してコネクタ83に入力されたシリアルデータは、外部接点bから共通接点cを介して入力側SERDES201に入力され、パラレルデータに変換される。このパラレルデータは、内部バス204を介して出力SERDES202に到達し、シリアルデータに変換される。このシリアルデータは、出力側スイッチ211の共通接点cから内部接点a及び導線213を介して、他方の出力側スイッチ212の内部接点aに入力される。そして、シリアルデータは、スイッチ212の内部接点aから共通接点cを介して、出力側SERDES202に入力され、パラレルデータに変換される。以下同様に、入力時の経路を逆に辿って、FC制御部31に戻る。

【0059】

このように、本実施例によれば、同一のディスク駆動部80に実装されるディスクドライブ群を複数に分割し、各ディスクドライブ群にそれぞれ異なるHDD制御基板82を割り当てるため、複数の障害が同時に発生した場合でも、障害に関連するHDD制御基板8

2のみを交換すればよく、障害発生時の耐性が向上する。また、障害耐性が高まることに伴い、失われたデータの回復処理等が行われる可能性も少なくなり、障害回復までの性能低下期間を短くすることができる。

【0060】

また、管理端末2から切替回路210にモード切替信号を出力し、追加されたDKA30にケーブル90を接続等するだけで、連結モードから独立モードへ簡単に移行させることができる。逆に、管理端末2から切替回路210にモード切替信号を出力し、DKA30へのケーブリングを廃止するだけで、独立モードから連結モードへ移行可能である。従って、簡単な操作でディスクアレイ装置10の運用モードを切替可能であり、使い勝手が向上する。さらに、各HDD制御基板82は、実質的に同一構成であるから、大量生産可能であり、ディスクアレイ装置10の製造コストを大幅に増大させることなく、柔軟な運用性を与えることができる。

【0061】

図11は、上述した第1の実施例の第1の変形例を示すブロック図である。本変形例では、各DKA30の各FC制御部31は、それぞれ複数のポートとデータ授受が行えるようになっている。即ち、各DKA30は、ディスク駆動部80と同数のFC制御部31をそれぞれ備えており、各FC制御部31は、ディスク駆動部80に設けられたディスクドライブ群の数に応じたポートのデータ処理を行うことができるようになっている。従って、この場合は、DKA30を追加することなく、連結モードから独立モードへ移行可能である。

【0062】

図12は、第2の変形例を示す説明図である。この変形例では、DKA30とディスク駆動部80、ディスク駆動部80同士のケーブリングをそれぞれ改良する。ディスク制御部11に実装されるDKA30は、各FC制御部31に対応するコネクタ31aをそれぞれ備えている。また、各ディスク駆動部80は、各HDD制御基板82に対応するコネクタ83をそれぞれ備えている。

【0063】

ここで、これら各コネクタ31a、83のうち、出力側のコネクタには、出力側コネクタであることを示す色彩（出力色）が施されている。また、入力側のコネクタには、入力側コネクタであることを示す色彩（入力色）が施されている。出力色としては、例えば灰色を用いることができ、入力色としては、例えば黒色を用いることができる。従って、ケーブル90が引き出されるDKA30のコネクタ31aには全て出力色（灰色）が付されている。HDD制御基板82につながるコネクタ83の場合は、DKA30または他のディスク駆動部80からのケーブル90が差し込まれる入力側のコネクタに入力色（黒色）が付され、他のディスク駆動部80へのケーブルを引き出すコネクタには出力色（灰色）が付されている。

【0064】

また、本変形例では、各ポート毎にケーブル90の色彩を変えている。即ち、図中上側に位置するAポートには、第1のポート色（黒色）が割り当てられており、図中下側に位置するBポートには、第2のポート色（青色）が割り当てられている。

【0065】

従って、保守作業員等は、配線手順書等を確認するまでもなく、コネクタの色彩とケーブルの色彩とによって、所定の機器同士を接続することができる。これにより、ディスクアレイ装置10の運用モードを変更したり、ディスク駆動部80を追加等する場合に、人為的な配線ミスが生じる可能性を低減することができる。

【0066】

なお、出力側コネクタ色及び入力側コネクタ色は、それぞれ単一の色彩である必要はなく、出力側コネクタ色及び／または入力側コネクタ色に複数の色彩を用いてもよい。ケーブル90の色彩も第1ポート色と第2ポート色に限らず、例えば、各コネクタ毎に異なる色彩を付与してもよい。

【0067】

図13は、第3変形例を示す説明図である。この変形例では、HDD制御基板82の接続回路230として、PBC (Port Bypass Circuit) を採用する。接続回路230は、出力側経路231と、入力側経路232と、入力側経路232の途中に設けられた複数のスイッチ233とを備えている。

【0068】

スイッチ233は、HDD制御基板82が担当するディスクドライブ81の数と同数だけ設けられる。各スイッチ233の接点bは、それぞれディスクドライブ81のSERDES81Aに接続されている。

【0069】

各ディスクドライブ81は、それぞれSERDES81Aを備えている。各SERDES81Aは、経路234を介して入力側経路232に接続されている。また、各SERDES81Aは、経路235を介してスイッチ233のb接点に接続されている。

【0070】

FC制御部31から出力されたシリアルデータは、コネクタ83からスイッチ212の外部接点b及び共通接点cを介して、入力側経路232に入力される。このシリアルデータは、経路234を介して各ディスクドライブ81のSERDES81Aに入力され、パラレルデータに変換される。

【0071】

ディスクドライブ81から読み出されたパラレルデータは、SERDES81Aによりシリアルデータに変換され、スイッチ233の接点b及び共通接点cを介して、入力側経路232に送られる。このシリアルデータは、入力側経路232から出力側のスイッチ212及びスイッチ211を介して、出力側経路231に送られる。さらに、このシリアルデータは、出力側経路231から入力側スイッチ211等を介して、FC制御部31に入力される。

【0072】

なお、本発明は、上述した実施の形態に限定されない。当業者であれば、本発明の範囲内で、種々の追加や変更等を行うことができる。例えば、同一のバックボードに実装されるディスクドライブ群は3個以上のグループに分割可能である。また、HDD制御基板82を連結する方法も上記の例に限らず、種々の方法を採用できる。例えば、内部接点同士をバックボードに形成されたプリント配線で接続する場合に限らず、外部接点同士をケーブル等で接続してもよい。この場合は、ケーブル接続のための手作業が必要となる。

【図面の簡単な説明】

【0073】

【図1】 本発明の実施例に係わるディスクアレイ装置を連結モードで運用する場合の全体概要を示すブロック図である。

【図2】 連結モードで運用されるディスクアレイ装置を正面から見た概略図である。

【図3】 DKAとディスク駆動部との接続関係を示すブロック図である。

【図4】 HDD制御基板の構造を示すブロック図である。

【図5】 複数のHDD制御基板を連結する場合のブロック図である。

【図6】 複数の障害が同時に発生した場合のブロック図である。

【図7】 ディスクアレイ装置を独立モードで運用する場合のブロック図である。

【図8】 独立モードで運用されるディスクアレイ装置を正面から見た概略図である。

【図9】 DKAとディスク駆動部との接続関係を示すブロック図である。

【図10】 HDD制御基板の構造を示すブロック図である。

【図11】 第1変形例に係わり、DKAとディスク駆動部との接続関係を示すブロック図である。

【図12】 第2変形例に係わるディスクアレイ装置のケーブリング方法を示す説明図である。

【図13】 第3変形例に係わるHDD制御基板の構造を示すブロック図である。

【図 1 4】従来技術によるディスクドライブと D K A との関係を示すブロック図である。

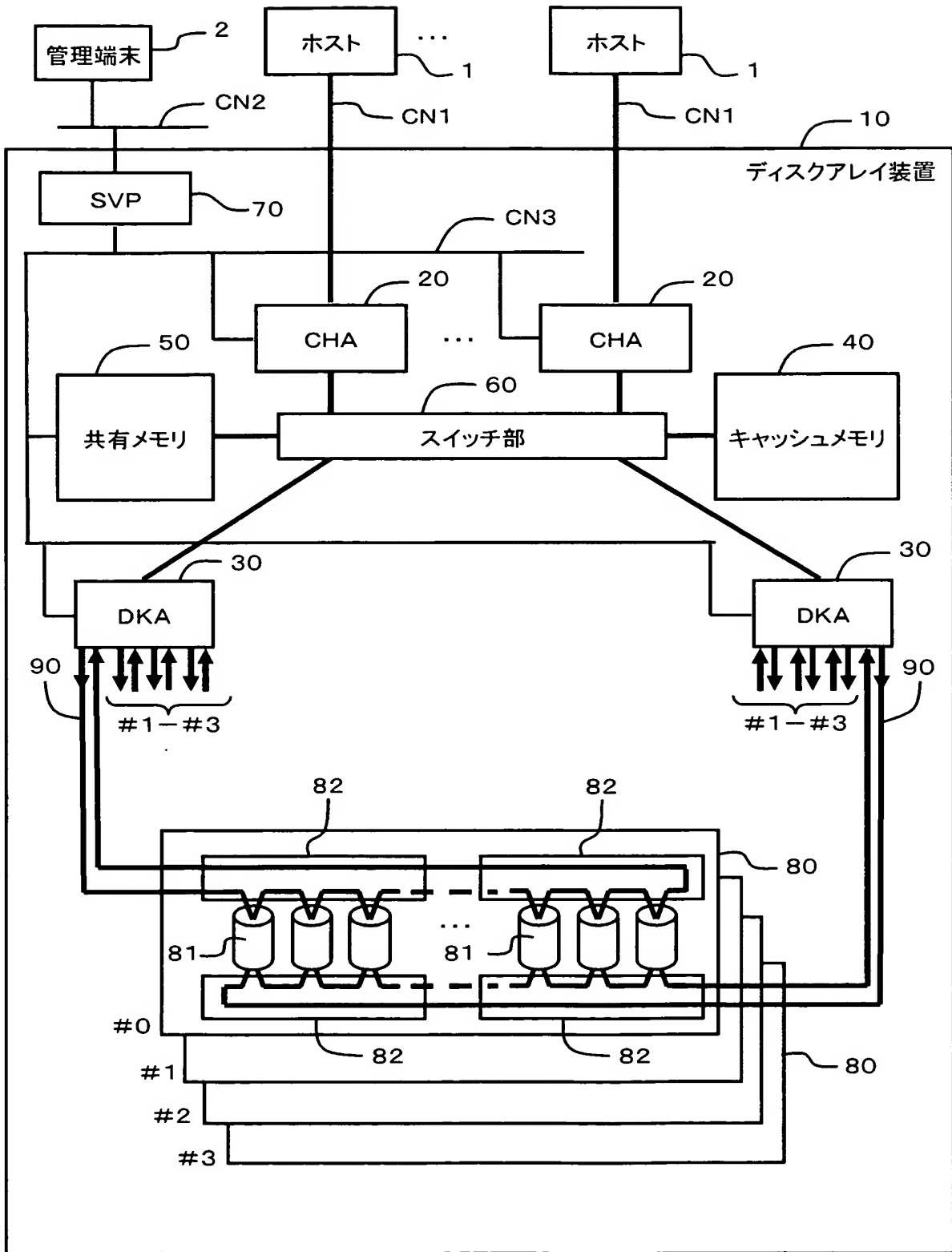
【図 1 5】他の従来技術によるディスクドライブと D K A との関係を示すブロック図である。

【符号の説明】

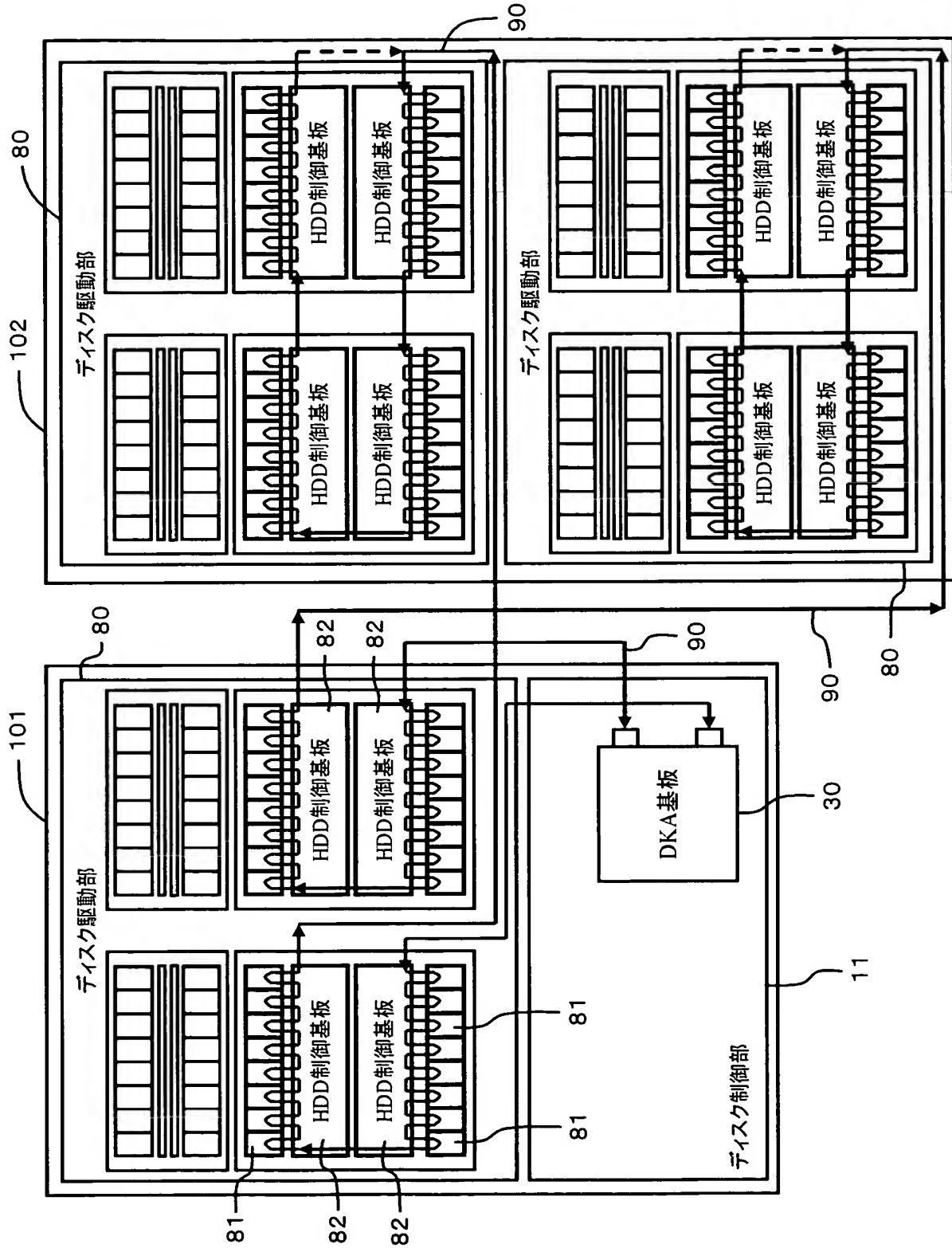
【0074】

1…ホストコンピュータ、2…管理端末、10…ディスクアレイ装置、11…ディスク制御部、20…C H A、30…D K A、31…F C 制御部、31 a…コネクタ、40…キャッシュメモリ、50…共有メモリ、60…スイッチ部、70…S V P、80…ディスク駆動部、81…ディスクドライブ、82…H D D 制御基板、83…コネクタ、90…ケーブル、101…基本部、102…増設部、200…接続回路、201～203…SERDES、204…内部バス、210…切替回路、211, 212…スイッチ、213…導線、230…接続回路、231…出力側経路、232…入力側経路、233…スイッチ、234, 235…経路、C N 1～C N 3…通信ネットワーク

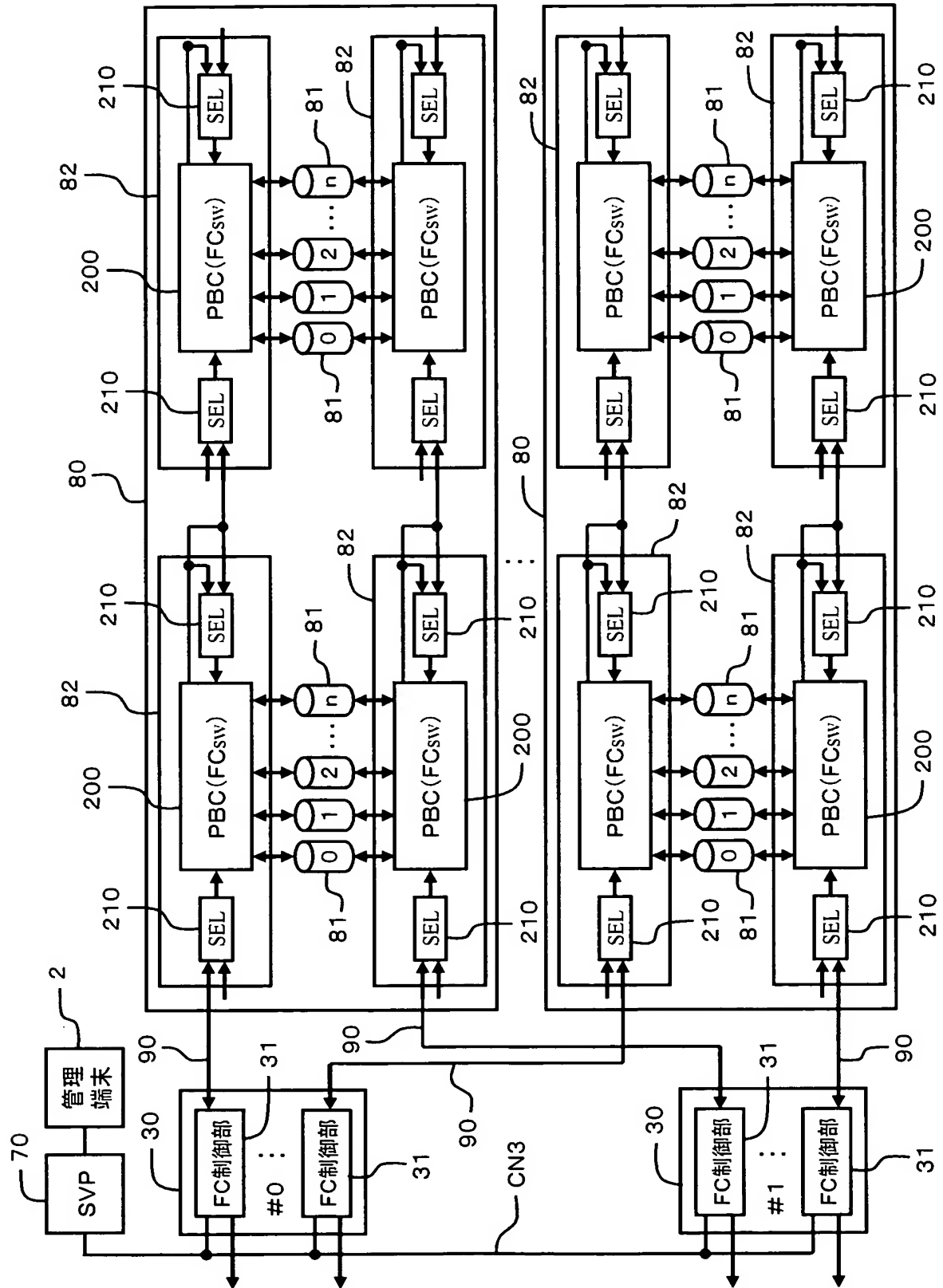
【書類名】 図面
【図 1】



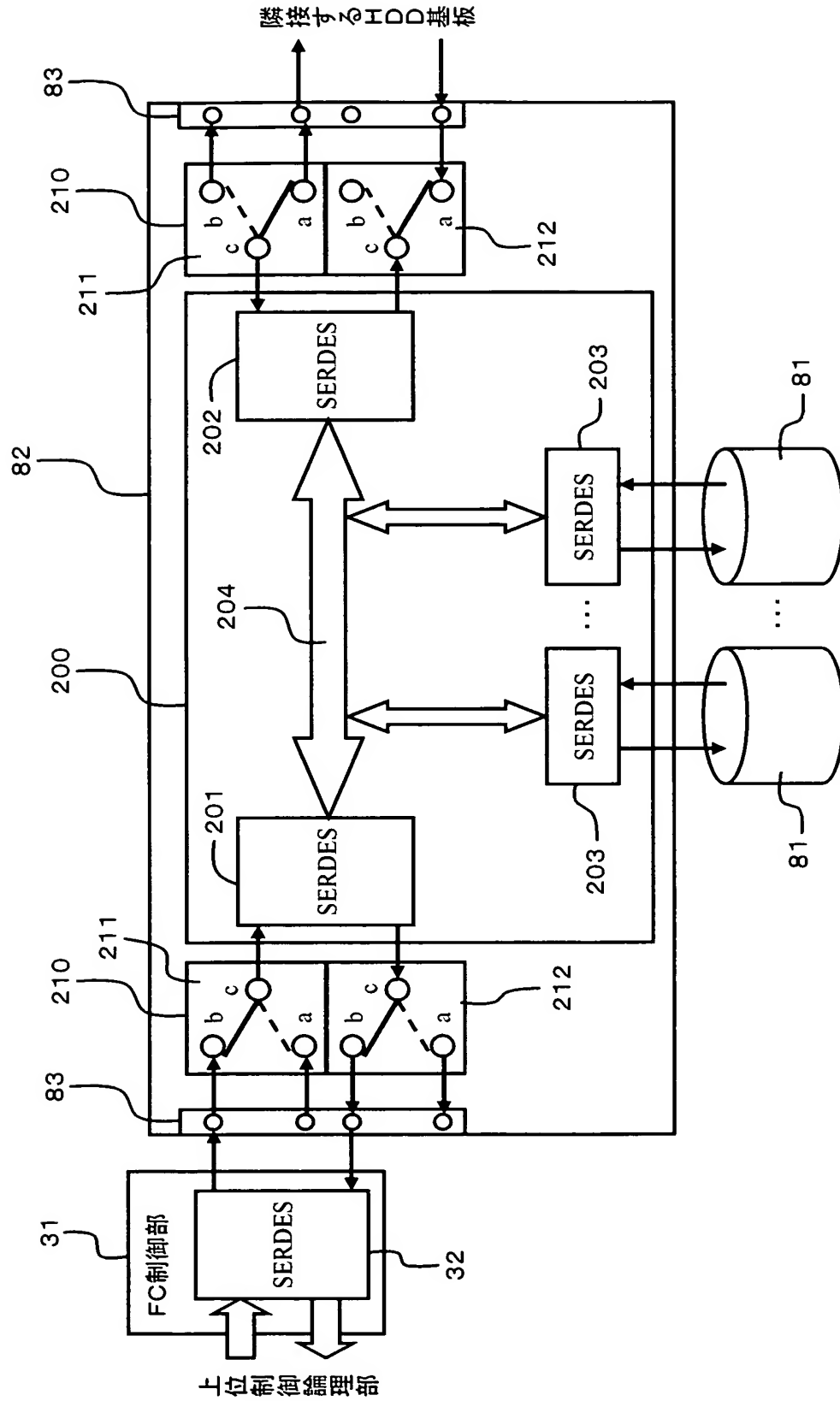
【図 2】



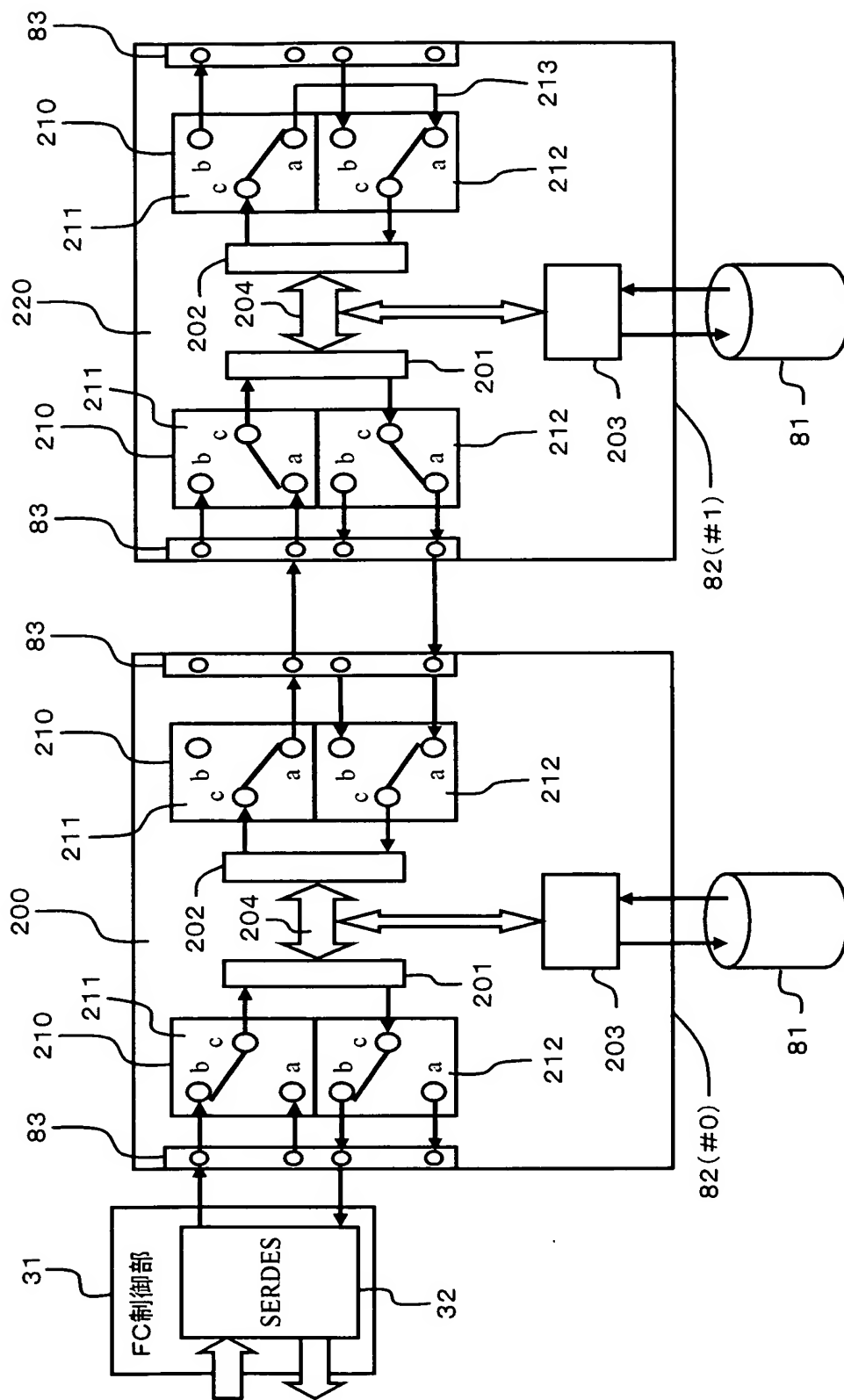
【図 3】



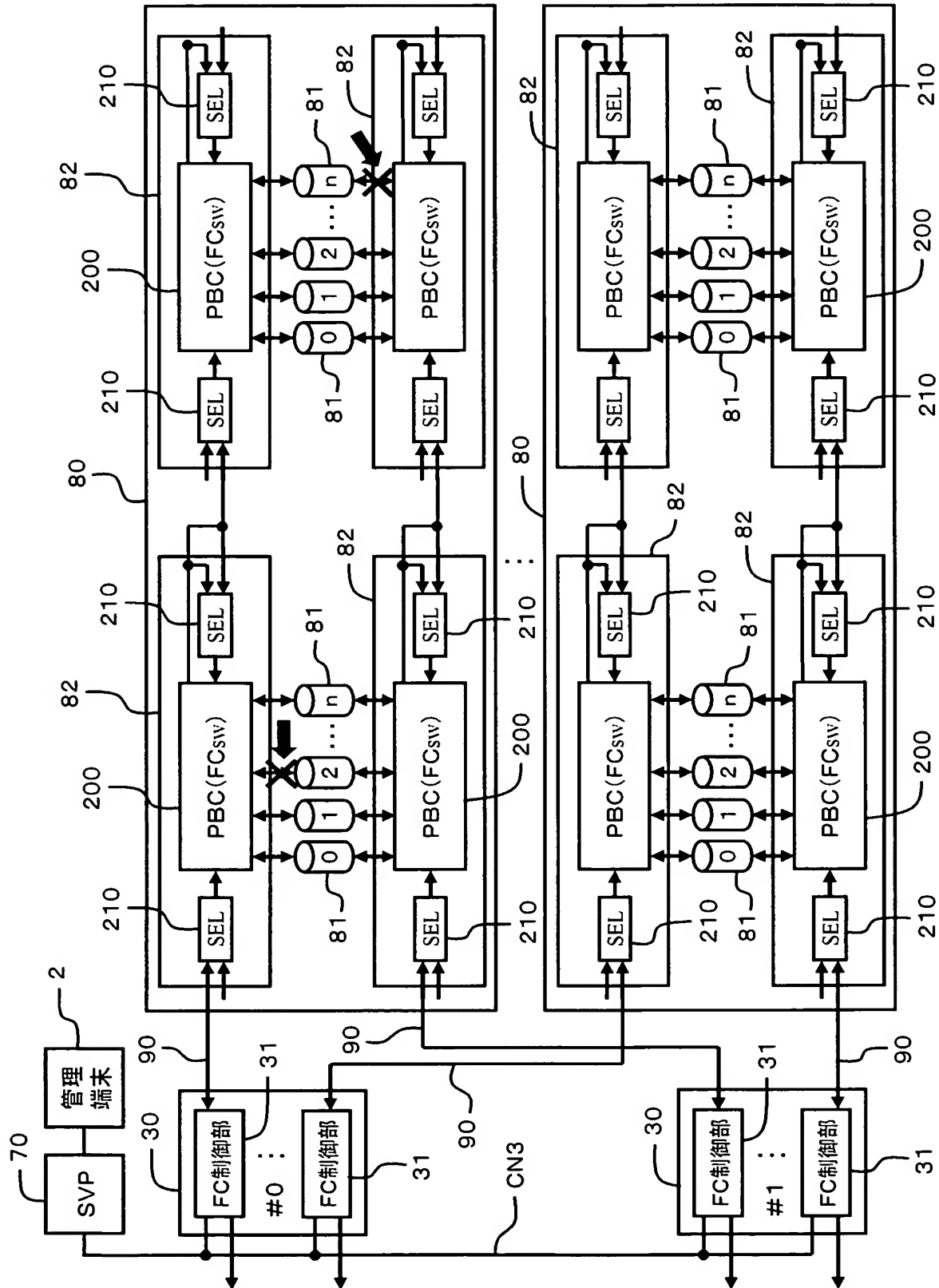
【図 4】



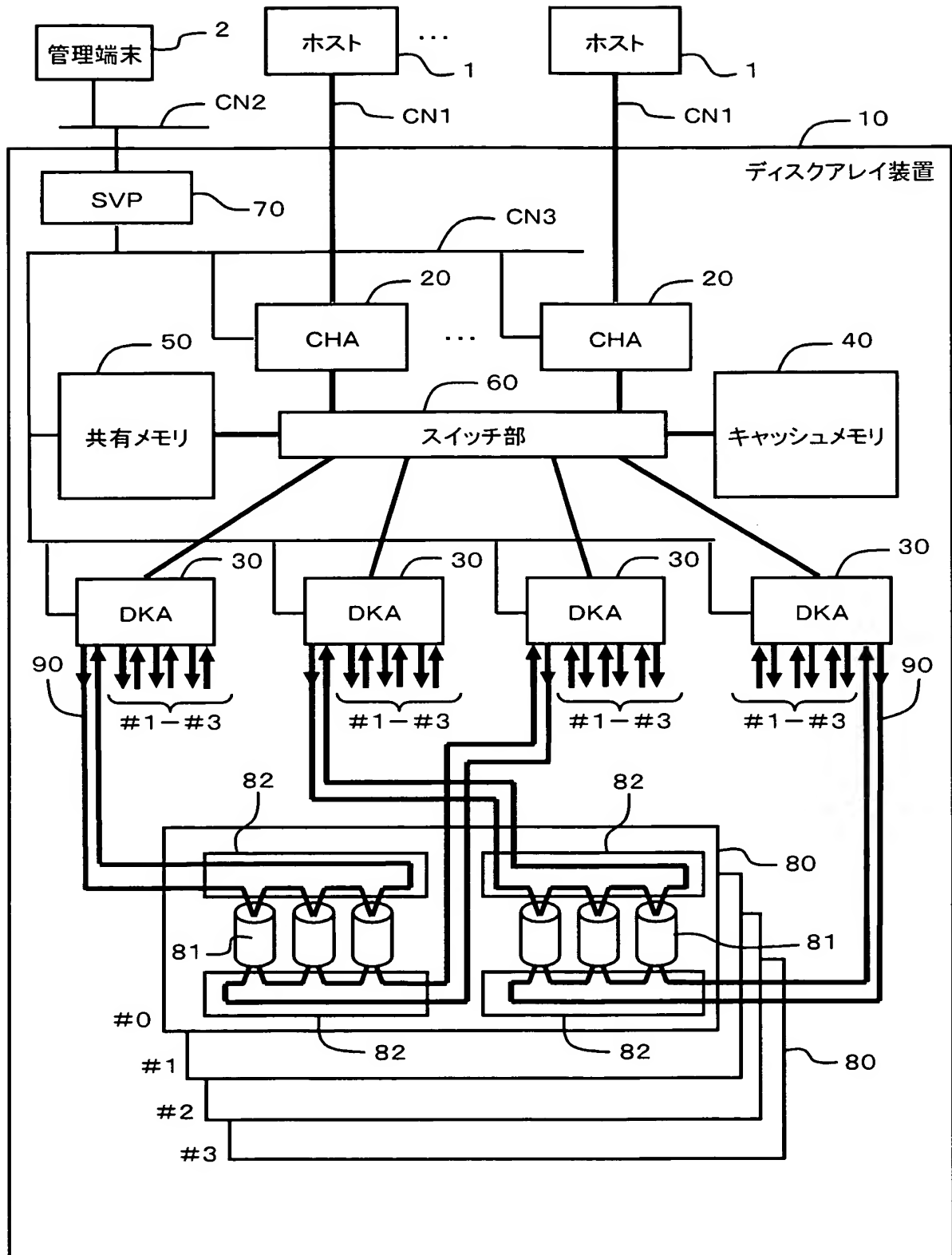
【図 5】



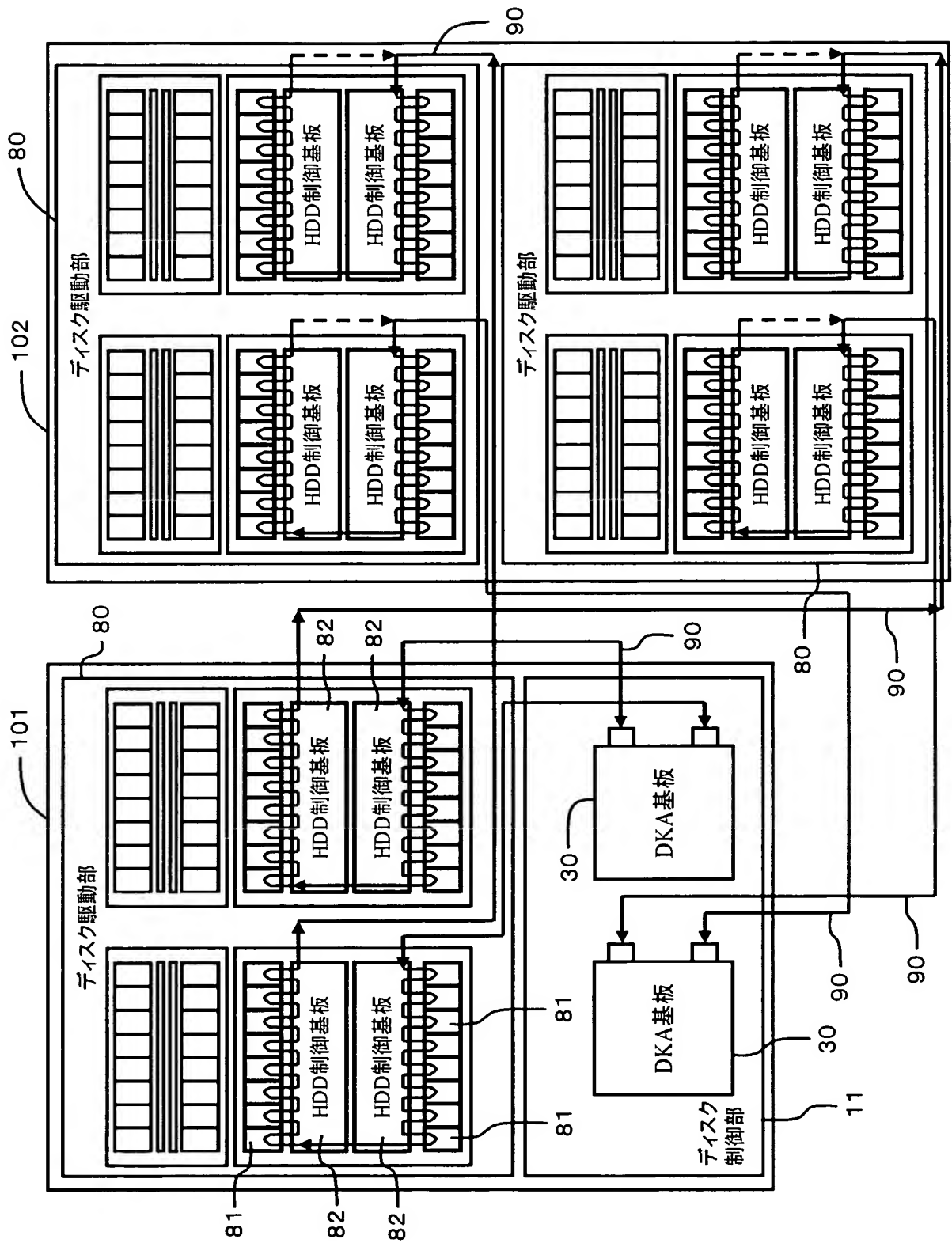
【図 6】



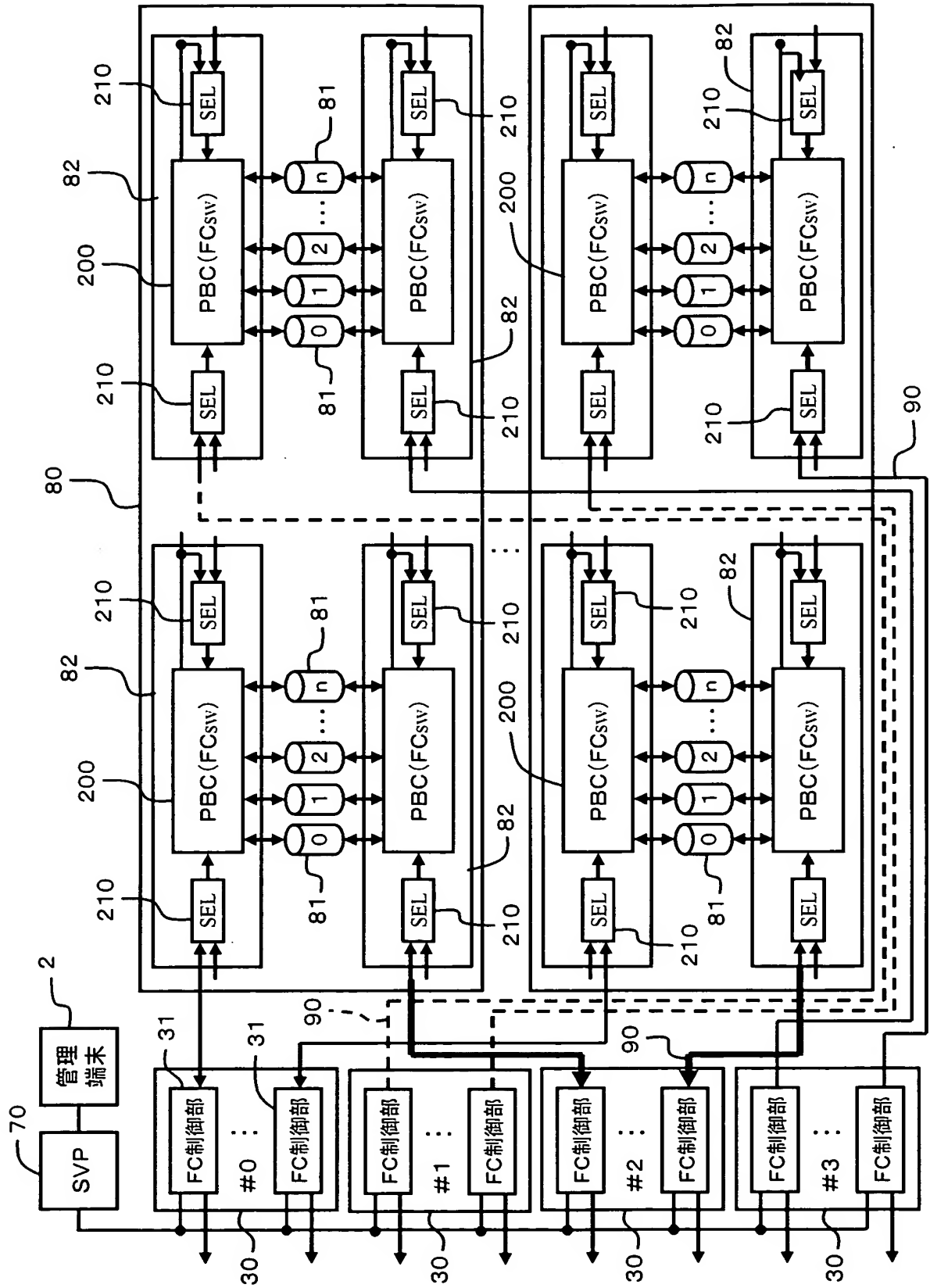
【図 7】



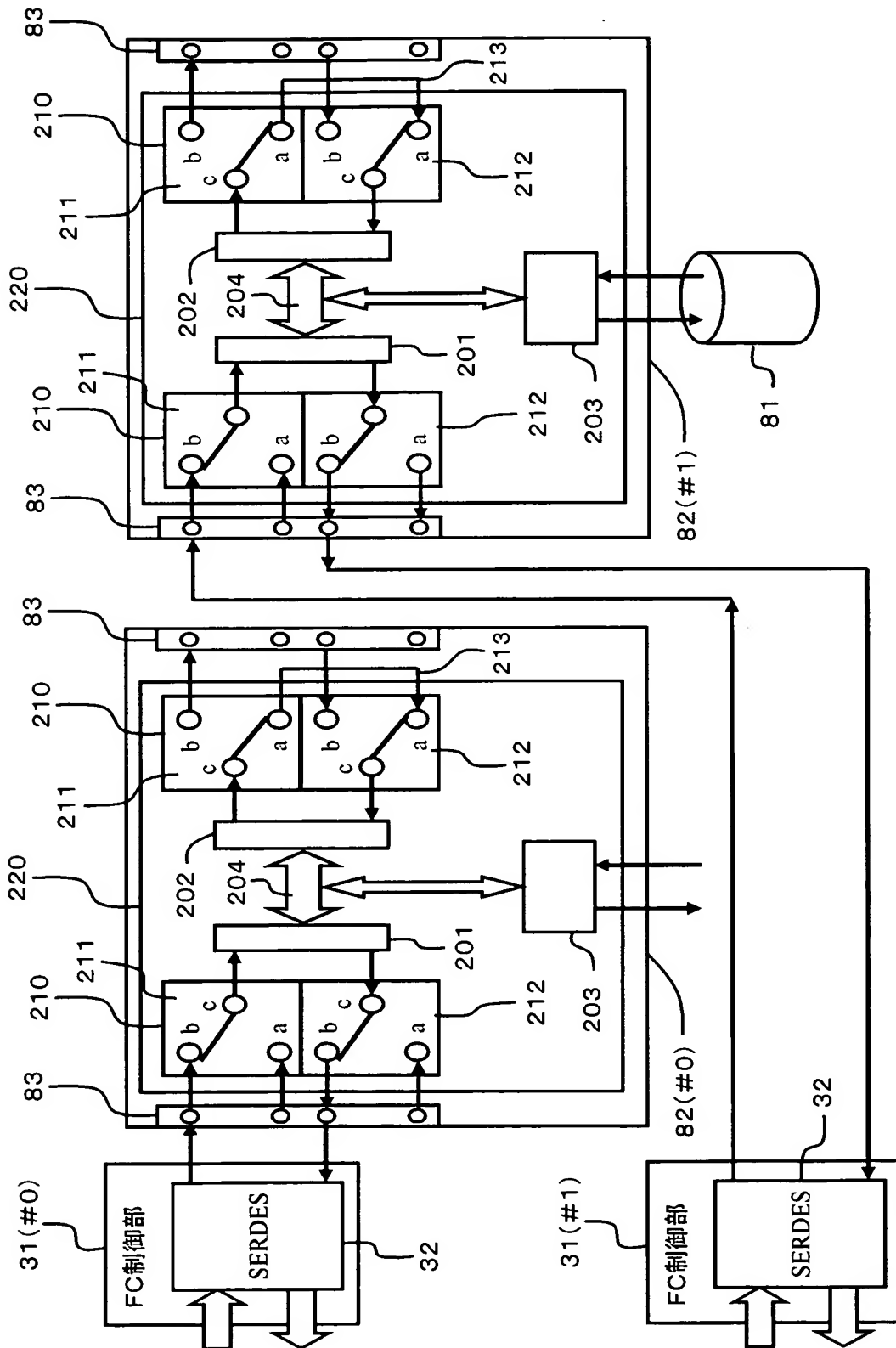
【図 8】



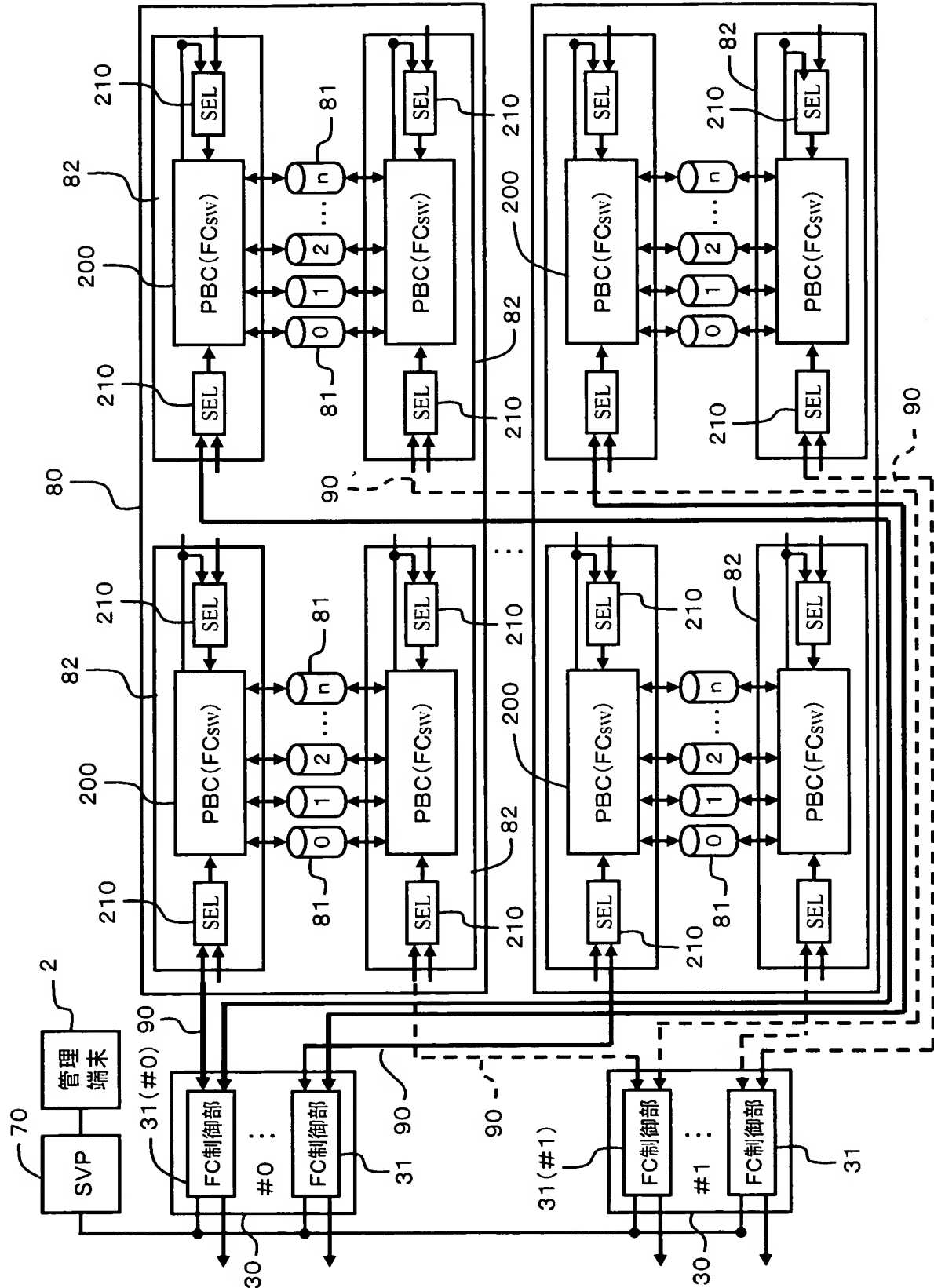
【図 9】



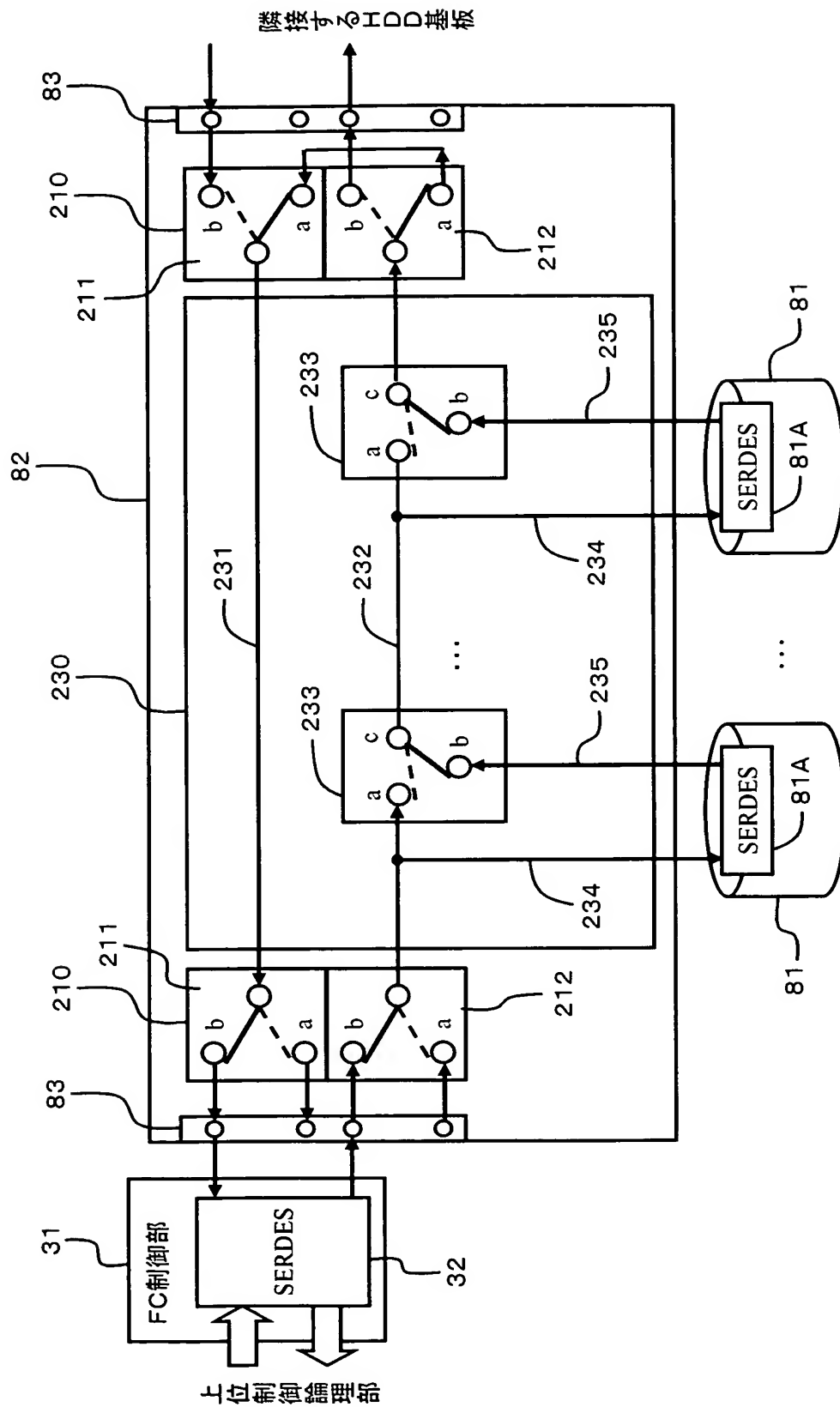
【図10】



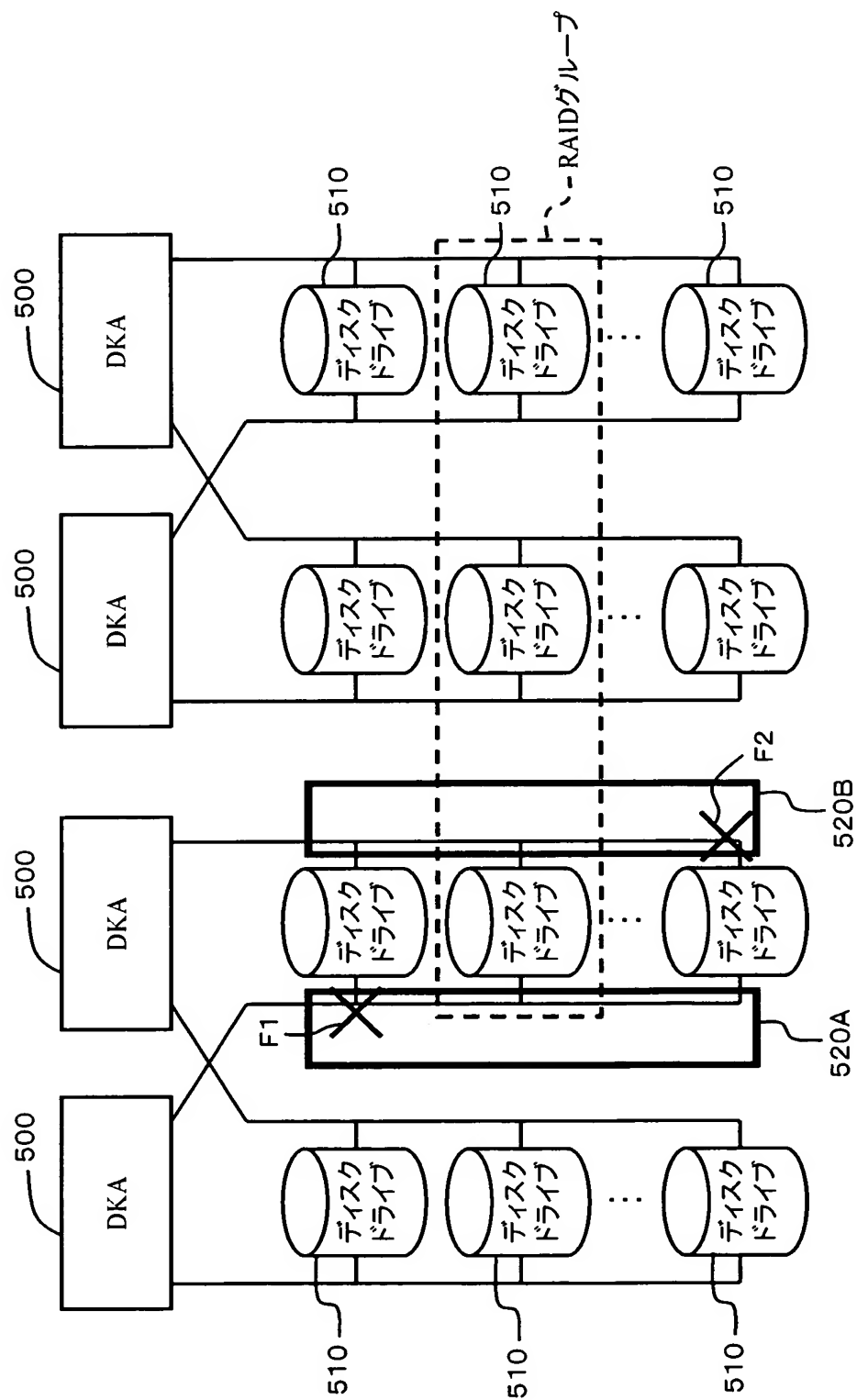
【圖 1 1】



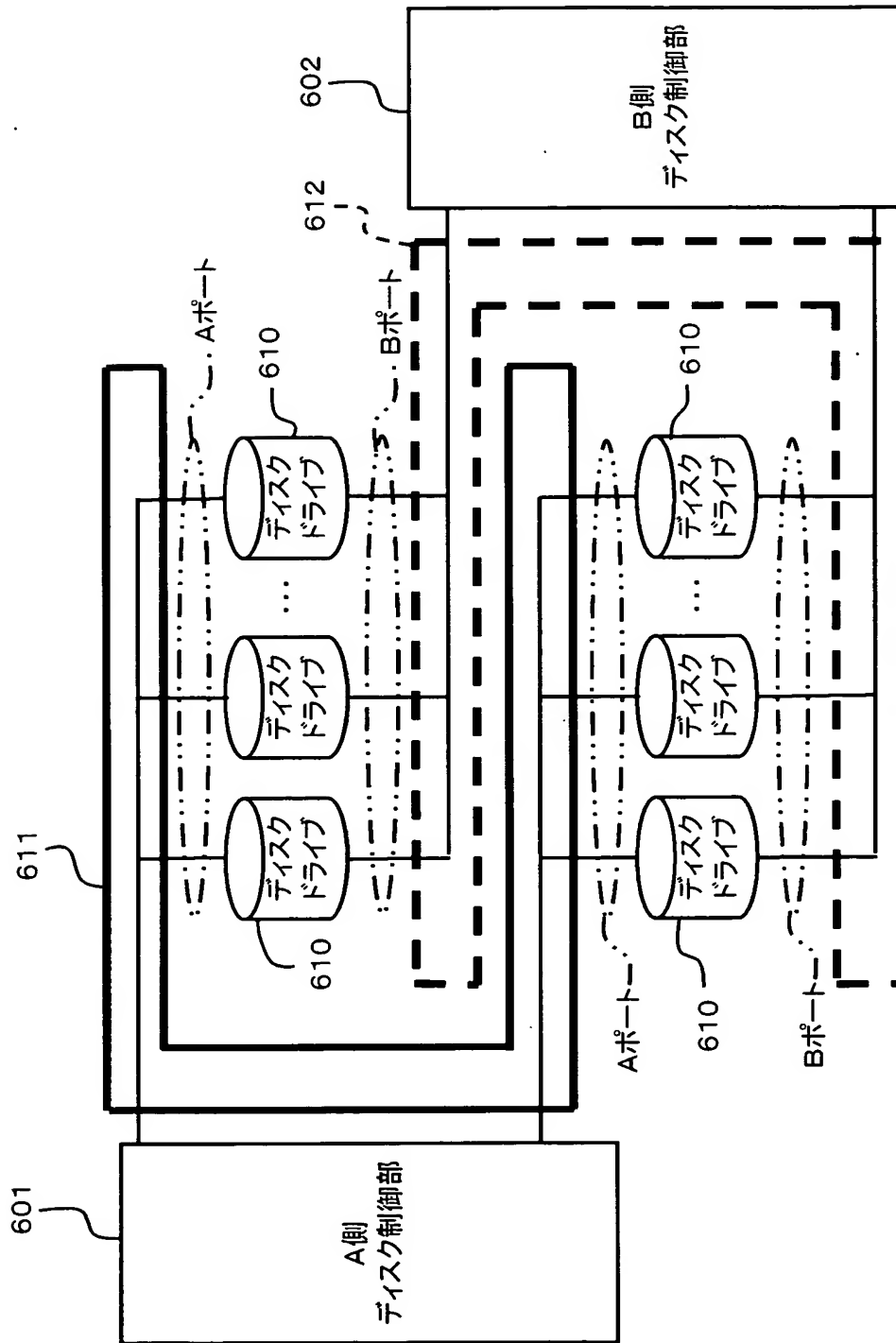
【図 13】



【図 14】



【図15】



【書類名】 要約書

【要約】

【課題】 ディスクアレイ装置の構成を使用目的等に応じて簡単に変更可能とする。

【解決手段】 同一のディスク駆動部には、それぞれ複数のディスクドライブ 81 から構成される複数のディスクドライブ群が設けられている。各ディスクドライブ群は、それぞれ別の HDD 制御基板 82 に接続される。HDD 制御基板 82 は、接続回路 200 と、切替回路 210 とを備える。管理端末から信号を出力して切替回路 210 を切り替えることにより、隣接する HDD 制御基板 82 同士を連結して運用することができる。また、切替回路 210 を切り替えることにより、隣接する HDD 制御基板 82 を互いに切り離し、独立して運用することができる。

【選択図】 図 5

認定・付加情報

特許出願の番号	特願 2 0 0 4 - 0 0 0 1 3 5
受付番号	5 0 4 0 0 0 0 1 5 9 5
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 6 年 1 月 6 日

< 認定情報・付加情報 >

【提出日】 平成16年 1月 5日

特願 2 0 0 4 - 0 0 0 1 3 5

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所